

PCT

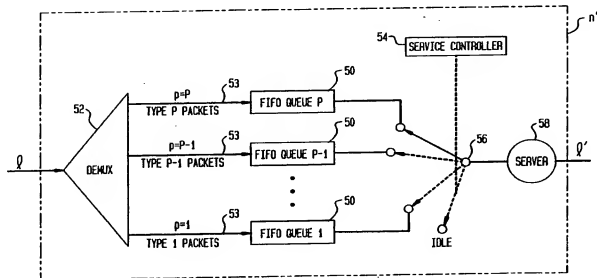
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification : H04J 3/24, H04Q 11/04	A1	(11) International Publication Number: WO 91/08628 (43) International Publication Date: 13 June 1991 (13.06.91)
<p>(21) International Application Number: PCT/US90/05621</p> <p>(22) International Filing Date: 3 October 1990 (03.10.90)</p> <p>(30) Priority data: 445,784 4 December 1989 (04.12.89) US</p> <p>(71) Applicant: BELL COMMUNICATIONS RESEARCH, INC. [US/US]; 290 West Mount Pleasant Avenue, Livingston, NJ 07039-2729 (US).</p> <p>(72) Inventor: GOLESTANI, S., Jamaloddin ; 114 Franklin Street, Apt. 3H2, Morristown, NJ 07060 (US).</p> <p>(74) Agents: WINTER, Richard, C.; PCT International, Inc., International Coordinator, Room 2E-304, Bell Communications Research, Inc., 290 West Mount Pleasant Avenue, Livingston, NJ 07039 (US) et al.</p>		<p>(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).</p> <p>Published With international search report.</p>

(54) Title: CONGESTION MANAGEMENT BASED ON MULTIPLE FRAMING STRATEGY



(57) Abstract

A congestion control strategy for a packet switching network (10) comprises an admission policy which controls the admission of packets into the network (10) and a stop-and-go queuing strategy (50) at the network nodes (n''). The congestion control strategy utilizes multiple frame sizes (Fig. 6A) so that certain types of connections can be provided with small queuing delays while other types of connections can be allocated bandwidth using small incremental bandwidth units.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CI	Côte d'Ivoire	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TO	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

CONGESTION MANAGEMENT BASED ON MULTIPLE FRAMING STRATEGY

Related Application

U.S. Patent Application Serial No. 326,027 entitled
5 "Congestion Free Packet Network" and filed on March 20,
1989 for S.J. Golestani is assigned to the assignee
hereof and contains subject matter related to the
subject matter of the present application. The above-
identified application is incorporated herein by
10 reference.

Field of the Invention

The present invention relates to a packet
transmission technique for alleviating congestion in a
15 packet switching network. More particularly, the
present invention relates to an improved congestion
control strategy which provides increased flexibility in
determining end-to-end packet delays and increased
flexibility in the allocation of transmission bandwidth
20 in accordance with the service requirements of different
classes of traffic.

Background of the Invention

The problem of congestion control, or more
25 generally traffic management, is quite significant for
packet switching networks. Congestion in a packet
switching network stems partly from the uncertainties
involved in the statistical behavior of many types of
traffic sources. Congestion is also due to the

complicated way in which different traffic streams interact with each other within a packet network. Congestion control and traffic management in a broadband integrated services environment is further complicated

5 by the high-speed of transmission and by the diverse mix of traffic types and service requirements encountered in such networks.

Most of the strategies proposed for congestion control in conventional narrowband data networks are

10 closed-loop in nature. This means that feedback information (e.g. acknowledgements) from the destination node or some intermediate nodes are used to decide about the admission of new packets into the network or about forwarding packets from one node to the next. At

15 broadband transmission rates, packet duration, or the time required by a link to serve a packet, is very short. Therefore, propagation delays, when measured in terms of packet duration, are much higher in broadband networks than in conventional narrowband networks.

20 Consequently, in a broadband network, any closed-loop or feedback-based control mechanism will tend to work more slowly and may be unable to keep up with the pace of events occurring in the network.

Services that may be encountered in a broadband

25 integrated services network range from data and voice communications to file transfers, high speed circuit emulations, and different types of video services. These services represent a wide variety of traffic

characteristics (e.g. average rate and burstiness) and a wide variety of service requirements (e.g. end-to-end delay, delay jitter, packet loss probability, call blocking probability, and error rate). The tasks of

5 resource management in general and congestion control in particular are more involved in this environment than in a conventional narrowband network. In a wideband integrated services network, control algorithms, besides having to deal with a wide range of traffic

10 characteristics, need to be effective in yielding predictable network behavior and must be flexible in accommodating different service requirements. Traffic management algorithms in a broadband network should also be simple in terms of the processing power required for

15 their execution. The increase in data processing speeds has not kept up with the fast growth of data transmission speeds. Thus, packet processing ability in the network nodes has more and more become the scarce resource. Therefore, the processing required for any

20 control function should be kept to a minimum.

The above-identified patent application discloses a congestion control strategy which has several highly desirable features: it maintains loss-free

communications, it provides bounded end-to-end delay,

25 and it is very simple to implement. These features make the strategy an attractive solution for the transmission of real time traffic and other forms of time-critical information in a broadband packet network.

The congestion control strategy of the above-identified patent application is composed of two parts: a packet admission policy imposed for each connection at its source node (i.e. a policy which controls the admission of packets into the network) and a service discipline at the switching nodes named stop-and-go queuing.

Central to both parts of the strategy is the notion of time frames. For this reason the congestion control strategy of the above-identified patent application is known as a framing strategy. On each link in the network, time intervals or frames of durations T are defined. The frames may be viewed as propagating from the transmitting end of a link to the receiving end of the link. Illustratively, the frames may be defined such that the frames on the incoming links at each node are synchronous or the frames on the outgoing links at each node are synchronous.

A stream of packets is defined to be (r, T) smooth when, for fixed length packets, the number of packets in each frame of duration T is bounded by $r \cdot T$, where r is a rate measured in packets/sec. The packet admission part of the congestion control strategy is based on the foregoing definition of smoothness. In particular, each connection k in the network has a source node and a destination node. The source node for connection k is the network node via which packets belonging to connection k enter into the network from an end user.

After a transmission rate r_k is allocated and reserved for a connection k along its path to its destination, the admission of packets to the network from this connection is required to be (r_k, T) smooth. This means

5 that during each frame of duration T , the source node for the connection k admits into the network no more than $r_k \cdot T$ packets and any additional packets are not admitted until the next frame starts. Alternatively, the allocated rate r_k may be required to be large enough

10 so that the stream of packets arriving at the source node and admitted to the network always maintains the (r_k, T) smoothness.

The above-identified packet admission policy guarantees that the traffic stream of connection k , with

15 an allocated rate r_k , is (r_k, T) smooth upon admission to the network. If this smoothness property continues to hold as the packet stream of each connection arrives at each intermediate switching node, then the problem of congestion is indeed resolved. Unfortunately, this is

20 not often the case. In a network which utilizes conventional first-in, first-out (FIFO) queuing at the nodes, as packets of a connection proceed from intermediate node to intermediate node, they tend to cluster together and form longer and longer bursts,

25 which bursts violate the original smoothness property.

The stop-and-go queuing technique is an alternative to conventional FIFO queuing which completely solves this problem. In particular, stop-and-go queuing

guarantees that once the (r_k, T) smoothness is enforced on all connections k at their source nodes, the property will continue to hold at any subsequent switching node. To facilitate discussion of the stop-and-go queuing
5 scheme, it is useful to consider arriving and departing frames. In particular, at each node, the arriving frames are the frames of duration T on the incoming links and the departing frames are the frames of duration T on the outgoing links. Over each link in the
10 network, the frames are viewed as traveling with the packets from one end of the link to the other. Therefore, when there is a propagation delay on link l and a processing delay at the receiving end of link l , which delays have a combined duration of r_l , the frames
15 at the receiving end of link l will be r_l seconds behind the frames at the transmitting end.

Accordingly, as indicated above, the frames may be defined so that the arriving frames on all incoming links of each node are synchronous with respect to each
20 other or so that the departing frames on all outgoing links of each node are synchronous with respect to each other. However, in general, at a node, the arriving and departing frames are asynchronous with respect to each other.

25 The stop and go queuing technique is based on the following rule: a packet which arrives at a node on an incoming link during an arriving frame f does not become eligible for transmission from the node until the

beginning of the first departing frame on the desired outgoing link which starts after f expires.

As a result of this rule, it follows that:

- 5 i) a packet that has arrived at some node via an incoming link during a particular arriving frame will be delayed and then be transmitted on the appropriate outgoing link in the first departing frame on the outgoing link which starts after the end of the particular
- 10 arriving frame on the incoming link. (Hence, the name stop-and-go queuing);
- ii) the packet stream of each connection k will maintain its original (r_k, T) smoothness throughout the network; and
- 15 iii) a buffer with a capacity of at most
$$B_l - 3C_l T \quad (1)$$
per link l is sufficient to eliminate any chance of buffer overflow, where C_l is the capacity of link l .

- 20 In a network which utilizes the above described framing congestion control strategy (including the above-described packet admission policy and the above-described stop-and-go queuing strategy), it is useful to consider the tradeoff between queuing delays and
- 25 flexibility in bandwidth allocation. More particularly, in the above-described congestion control strategy, the

combined queuing delay at all nodes for each connection is given by

$$Q = \alpha HT \quad (2)$$

where α is a constant between 1 and 2 and H is the
5 number of links in the connection (e.g. a connection which utilizes two links has one intermediate node). Equations 1 and 2 indicate that by choosing a sufficiently small frame size T, one can arbitrarily reduce queuing delays as well as buffer requirements in
10 the network.

In addition, for fixed length packets of L bits in length, the incremental unit of bandwidth allocation is

$$\Delta r = L/T \text{ bits/sec} \quad (3)$$

15 Equation (3) indicates that a small frame size T (and the resulting small buffer requirements and queuing delays) can only be achieved at the expense of a large
20 incremental unit of bandwidth allocation and thus decreased flexibility in bandwidth allocation.

Briefly stated, it is a shortcoming of the above-described framing congestion control strategy that small queuing delays and small buffer capacity requirements
25 can only be achieved at the expense of decreased bandwidth allocation flexibility.

Accordingly, it is an object of the present invention to provide a modification of the above-described framing congestion control strategy so that it
30 is possible to enjoy small queuing delays for certain

connections while still being able to allocate bandwidth using arbitrarily small incremental bandwidth units for other connections.

It is a further object of the present invention to
5 provide a framing congestion control strategy which utilizes multiple frame sizes so that some connection can enjoy small queuing delays while other connections can be allocated bandwidth in small incremental units.

10 Summary of the Invention

In accordance with the framing congestion control strategy of the present invention, multiple frame types with multiple durations are utilized. In particular, the frame sizes or durations are $T_1, T_2 \dots T_p \dots T_p$
15 wherein T_p is a multiple of T_{p+1} so that $T_p = M_p T_{p+1}$, where M_p is an integer. On each link in the network time frames of duration T_p are defined. This is repeated for every value of p , illustratively, starting from the same time origin for each link.
20 Illustratively, $P=2$ and $T_1=3T_2$. The frames of multiple sizes on each link may be viewed as traveling from the transmitting end to the receiving end of each link. Since, on each link, the multiple frame types have durations which are integrally related and are defined
25 from the same time origin, the multiple frame types on each link are synchronous with respect to each other. The multiple frame types may be defined so that at each node all the arriving frames on the incoming links are

synchronous with respect to each other or so that at each node all the departing frames on the outgoing links are synchronous with respect to each other. In general, at a node, arriving and departing frames may be
5 asynchronous with respect to each other.

Every connection in the network is set up as a type- p connection for some $p=1,2,\dots,P$, in which case the connection is associated with the frame size T_p .

The multiple frame size congestion control strategy
10 of the present invention involves an admission policy at the source node of each connection, and a stop-and-go queuing technique at each node.

The admission policy requires that the packet arrival of any connection k of type- p with the allocated
15 and reserved transmission rate r_k be (r_k, T_p) smooth. Packets of a type- p connection are referred to as type- p packets, which packet type is indicated in the packet header.

The stop-and-go queuing strategy at the network
20 nodes utilizes the following rules:

- A) A type- p packet which has arrived at a node on an incoming link during an arriving type- p frame f , does not become eligible for transmission from the node until the beginning
25 of the first type- p departing frame on the appropriate outgoing link which starts after f has expired.

B) Any eligible type-p packet has non-preemptive priority over a type-p-1 packet. This means that a higher priority type-p packet which becomes eligible for service by an outgoing link of a node while a lower priority type-p-1 packet is being served, waits until service of the type-p-1 packet is completed before gaining priority over the outgoing link.

As a result of these rules, it follows that:

(i) any type-p packet that has arrived at a node via an incoming link during an arriving type-p frame f , will receive service on an outgoing link during the first departing type-p frame which starts after the incoming type-p frame f has ended;

(ii) the packet stream of each connection will maintain its original smoothness property along its entire path through the network, i.e., for a type-p connection with an allocated transmission rate r_k , the packet stream maintains the $(r_k.T_p)$ smoothness throughout its path to its destination;

(iii) a buffer capacity of

$$B_i = 3 \sum C_i^p T_p \quad (4)$$

per outgoing link i is sufficient to eliminate any chance of buffer overflow, where C_i^p is the capacity of link i allocated to type-p

connections. This buffer capacity is always less than $3C_gT_1$.

It is a significant advantage that the multiple frame size congestion control strategy enables certain
5 connections to have very small queuing delays, while for other connections, bandwidth can be allocated in small incremental bandwidth units.

More particularly, the queuing delay Q^p for a type-p connection is given by

$$10 \quad H \cdot T_p < Q^p < 2H \cdot T_p \quad (5)$$

where H is the number of links of the connection. Thus small frame sizes are utilized for connections which can tolerate only small queuing delays.

With respect to bandwidth allocation, the
15 incremental bandwidth unit for a type-p connection is

$$\Delta r^p = L / T_p \text{ bits/sec} \quad (6)$$

where L is the packet length. Thus for connections which require that bandwidth be allocated in small incremental units, large frame sizes are utilized.

20 In short, in the multiple frame size congestion control strategy, frame size and queuing delay are coupled together for each connection type individually, rather than globally as is the case for the single frame size case, thereby providing a significant increase in
25 the flexibility of the congestion control strategy.

Brief Description of the Drawings

FIG 1 schematically illustrates a packet switching network and the path of a connection in the network.

FIG 2A schematically illustrates a packet stream which is (r, T) smooth for $r-2/T$ and FIG 2B schematically illustrates a packet stream which is not smooth for $r-2/T$.

FIGs 3A and 3B illustrate alternative implementations of a network node for the framing congestion control strategy when there is a single frame size.

FIGs 4A and 4B illustrate arriving and departing frames for the nodes of FIGS 3A and 3B when the arriving and departing frames are asynchronous and synchronous, respectively.

FIG 5 compares queuing delay distributions for stop-and-go queuing and conventional FIFO queuing.

FIG 6A illustrates multiple frame sizes for use in an illustrative embodiment of the congestion control strategy of the present invention.

FIG 6B illustrates frames of multiple sizes at the transmitting end of a link l , at the receiving end of the link l at a node n^* , and at the transmitting end of the link l' at the node n^* .

FIG 7 illustrates a network node which is capable of carrying out the multiple frame size congestion control strategy of the present invention.

FIG 8 illustrates a typical pattern of service shifts between queues in the node of FIG 7.

FIG 9 illustrates an alternative node in which stop-and-go queuing is combined with another queuing discipline.

Detailed Description of the Invention

Before describing the present invention in detail, it may be helpful to briefly describe the structure of a packet switching network and the definition of smoothness utilized in the framing congestion control strategy with both single and multiple frame size.

FIG 1 schematically illustrates a packet switching network 10. The network 10 comprises the nodes $n-1, 2, 3, 4$ and 5 and the links $l-1, 2, 3, 4$. Illustratively, two connections are shown in FIG 1. The connection $k-1$ extends from the source node $n-1$, through the intermediate nodes $n-2$ and $n-3$, and to the destination node $n-4$. Thus, the connection $k-1$ utilizes the links $l-1, 2, 3$. Similarly, the connection $k-2$ extends from the source node $n-4$, through the intermediate node $n-2$, and to the destination node $n-3$. Thus, the connection $k-2$ utilizes the links $l-4$ and $l-2$. Each of the connections k in the network is allocated a rate r_k in terms of the number of packets per second.

A packet stream belonging to the connection k and propagating on one of the links l is defined as (r_k, T) smooth if the number of packets in a time frame T does

not exceed $r_k \cdot T$. The packet stream is not smooth if the packets are clustered so that for some of the frames the number of packets exceeds $r_k \cdot T$.

FIG 2A shows a sequence of frames of duration T on a link l . Each packet being transmitted via the link l is designated with an arrow. The packet stream of FIG 2A is $(2/T, T)$ smooth since the number of packets in a frame never exceeds $(2/T) \cdot T - 2$. In contrast, the packet stream of FIG 2B is not $(2/T, T)$ smooth since in the frame $[T, 2T]$ and the frame $[4T, 5T]$ the packets are clustered so that the number of packets (i.e. 3) exceeds $(2/T) \cdot T - 2$.

As indicated previously, the above mentioned U.S. patent application describes a congestion control strategy for a packet network which comprises a packet admission policy imposed for each connection at the source node, and a queuing strategy to be carried out at the nodes known as stop and go queuing.

The packet admission part of the congestion control strategy is based on the foregoing definition of smoothness. After a transmission rate r_k (in packets per second) is allocated to a connection k , the admission of packets for that connection into the network is required to be (r_k, T) smooth, i.e., during each frame, once the source node for connection k receives $r_k \cdot T$ packets no more packets are admitted to the network for connection k until the next frame starts. Preferably, the allocated rate r_k for a

connection k is large enough so that the stream of packets entering the network always maintains the (r_k, T) smoothness.

FIG 3A schematically illustrates a node n for
5 implementing the stop-and-go queuing strategy. FIG 3B shows an alternative node n' for implementing the stop-and-go queuing strategy. In FIGs 3A and 3B, packets arrive at the node n or n' on an incoming link l and packets depart the node n or n' on an outgoing link l' .
10 Typically, a node n or n' has a plurality of incoming links l and a plurality of outgoing links l' . However, for purposes of clarity, only one incoming link and one outgoing link are shown for the nodes n and n' .

The arriving frames on the incoming link l and the
15 departing frames on the outgoing link l' may be asynchronous or synchronous with respect to each other. Asynchronous arriving and departing frames of duration T on the links l and l' , respectively, are shown in FIG 4A. Synchronous arriving and departing frames of
20 duration T on the links l and l' , respectively, are shown in FIG 4B.

The stop-and-go queuing rule to be implemented at the nodes may be stated as follows. A packet which arrives at a node on an incoming link l in an arriving
25 frame f is queued at the node and served by the outgoing link l' in the first departing frame f' which starts after the end of the arriving frame f . Examples of

arriving and departing frames f and f' are shown in FIGS 4A and 4B for the asynchronous and synchronous cases.

FIGS 3A and 3B illustrate node structures n and n' for implementing the stop-and-go queuing strategy in the case where the arriving and departing frames are synchronous.

In FIG 3A the node n comprises two FIFO queues 12 and 14 and two switches 16 and 18. Thus, during a first frame, the incoming link l is connected by the switch 16 to the FIFO queue 12 in which arriving packets are stored. During this same first frame, the switch 18 is connected to the FIFO queue 14 so that packets arriving via the link l in the previous frame are served by the server 19 and transmitted via the outgoing link l' . In the next frame, the positions of the switches 16 and 18 are reversed, so that arriving packets are stored in the queue 14 and packets which were stored in the queue 12 during the first frame are served by the server 19 and transmitted via the outgoing link l' . The arrangement of FIG 3A insures that packets arriving at the node n via the link l during a particular frame are not served by the server 19 and transmitted via the outgoing link l' until the next frame.

The node n' of FIG 3B accomplishes the same thing using only a single FIFO queue 20. In the node n' , at the beginning of each frame, a service controller 22 marks the present load in the queue 22 as eligible. The controller 22 then connects server 19 to the queue 20

via the switch 24 for just the time needed to serve the eligible load. After that, the position of the switch 24 is changed to interrupt service of the queue 20 by the server 19 until the beginning of the next frame so that any packet which is received during the current frame, and therefore not yet eligible for transmission, does not get transmitted before the expiration of the current frame. A modification of the node n' of FIG 3B for the case where the arriving and departing frames are asynchronous is discussed in the above-mentioned U.S. Patent application, Ser. No. 326,027.

It should be noted that stop-and-go queuing does not require packets to be served on a FIFO basis. Nevertheless, in particular implementations of stop-and-go queuing nodes, such as the nodes n and n' of FIGS 3A and 3B, the packets are served on a FIFO basis. This is advantageous because it simplifies implementation of the nodes and insures that packets are delivered in proper order.

As indicated above, the framing congestion control strategy has a number of significant advantages. First, when the admissions policy at the source nodes and the stop-and-go queuing strategy are utilized, the required (r_k, T) smoothness is maintained for each connection in the network over all the links traversed by the congestion. This means that the framing congestion control strategy prevents the clustering of packets, which clustering is a cause of congestion. In addition,

a buffer capacity of $2C_2T$ for synchronous arriving and departing frames and at most $3C_2T$ for asynchronous arriving and departing frames is sufficient to eliminate any chance of buffer overflow.

5 The foregoing packet admission and queuing strategy, besides eliminating buffer overflow, also leads to an attractive property in terms of packet delays. All packets of any given connection undergo the same amount of delay in the network, except for a small
10 jitter. In other words, the end-to-end delay, D' , of a packet in a given connection may be represented as

$$D' = D + d \quad (7)$$

where D , defined as the connection delay, is constant for all packets of the connection, and d , the delay jitter, varies from packet to packet. The delay jitter is limited to the size of a frame:

$$-T < d < T \quad (8)$$

The connection delay D can be broken into two parts:

$$20 \qquad D = t + Q \qquad (9)$$

where τ is the sum of the end-to-end propagation and processing delays of a connection and Q represents the total queuing delays of the connection. Furthermore, Q is bounded as follows:

25 Q - aHT (10)

where H is the number of links traversed by the connection and α is a constant between 1 and 2.

FIG 5 compares the packet delay distribution for conventional FIFO queuing with the packet delay distribution of stop-and-go queuing. Thus, in FIG 5, curve 30 is the distribution of packet delays resulting from stop-and-go queuing and curve 40 is the distribution of packet delays resulting from conventional FIFO queuing. The shaded region 42 roughly corresponds to the fraction of packets that are lost when the conventional FIFO queuing is utilized. While conventional FIFO queuing has a smaller average delay, the delays are distributed over a wide range, which makes the network subject to buffer overflow and packet loss. In contrast, the delay performance of the framing congestion control strategy is within a narrow but acceptable range so that the behavior of the network is predictable and satisfactory.

As indicated above, both the required buffer capacity at each link and the queuing delay Q can be made arbitrarily small, by choosing a sufficiently small frame size T . However, a small frame size T can only be chosen at the cost of reduced flexibility in bandwidth allocation. In particular, the unit of bandwidth allocation is given by

$$\Delta r = L/T \text{ bits/sec} \quad (11)$$

so that as T decreases Δr increases. As a numerical example, for a packet length of $L=400$ bits and a connection with $M=5$ links, and for $\alpha = 1.6$, the queuing delay Q is $1.6HT$. If $T=1\text{msec}$, the queuing delay is

8msec and Δr is 400 kbits/sec. On the other hand, if a bigger T is chosen (e.g. $T=50$ msec) to reduce Δr to 8 kbits/sec, the queuing delay is 400 msec.

- This problem is alleviated in accordance with the present invention by utilizing multiple frame sizes. Thus a plurality of frame sizes T_p are chosen with $p=1,2,\dots,P$. T_p is a multiple of T_{p+1} , i.e., $T_p = M_p T_{p+1}$, where M_p is an integer. Thus, on each link of the network, time frames of duration T_p are defined. This is repeated for every value of p , illustratively, starting from the same time origin on each link. Since, on each link, the multiple frame types have durations which are integrally related and are defined from the same time origin, the multiple frame types on each link are synchronous with respect to each other. The multiple frame types may be defined so that at each node all the arriving frames on all the incoming links are synchronous with respect to each other, or so that at each node all the departing frames on all the outgoing links are synchronous with respect to each other. In general, at a node, arriving and departing frames may be asynchronous or synchronous with respect to each other. FIG 6A illustrates type-1 frames and type-2 frames on a link for the case where $T_1=3T_2$. Note that in FIG 6A, the type-1 and -2 frames are synchronous with each other. This occurs because the type-1 and type-2 frames are defined from the same time origin and because T_1 is an integer multiple of T_2 .

FIG 6B shows a node n^* and an incoming link l and a departing link l' . FIG 6B shows type-p-1 and type-p-2 frames 200 at the transmitting end 90 of link l . FIG 6B also shows the type-p-1 and type-p-2 frame 210 at the receiving end 92 of the link l at the node n^* . Note, that the propagation delay on link l and the processing delay at the receiving end of link l have a combined duration of r_l . Thus the frames at the receiving end of the link l are delayed r_l seconds with respect to frames at the transmitting end of the link l .

FIG 6B also shows the type-p-1 and type-2 frames 220 departing the node n^* at the transmitting end 94 of the link l' . Note that the arriving frames 210 and departing frame 220 at the node n^* are asynchronous. Thus, θ_l^2 represents the amount of delay which has to be added to the link l for type-2 packets in order to synchronize the arriving and departing type-p-2 frames at the node n^* . Similarly, θ_l^1 represents the amount of delay which has to be added to the link l for the type-1 packets in order to synchronize the arriving and departing type-p-1 frames at the node n^* .

As indicated above, the inventive multiple frame size congestion control strategy involves an admissions policy at the source node of each connection and a queuing strategy at the network nodes.

The admissions policy is as follows. Every connection in the network is set up as a type-p connection. The admissions policy requires that the

packet admission into the network at the source node of any connection k of type- p with the allocated transmission rate r_k , be $(r_k.T_p)$ smooth. A packet belonging to a type- p connection is a type- p packet, which packet type is indicated in the packet header.

To prevent clustering of packets and the congestion that results therefrom, the stop-and-go queuing strategy at the nodes is based on the following rules.

- A) A type- p packet which has arrived at a node via an incoming link i in an arriving type- p frame f does not become eligible for service until the beginning of the first type- p departing frame on the appropriate outgoing link which starts after f has expired.
- B) Any eligible type- p packet has non-preemptive priority over type $p-1$ packets, e.g. type-2 packets have non-preemptive priority over type-1 packets.

Generally, capacity of the links is allocated such that the aggregate transmission rate allocated over link i to all type- p connections, denoted by C_i^p (in bits/sec) satisfies the following:

$$\sum_{p=p_0}^P C_i^p \leq C_i - L/T_{p_0}, \quad p_0=2, \dots, P \quad (12)$$

The term L/T_{p_0} in this inequality is a rate in bits/sec equivalent to the transmission of one packet per frame of duration T_{p_0} . Therefore, this inequality suggests

that for each link l and for any $p_0 \geq 2$, at least one packet per frame of duration T_{p_0} of the link capacity should remain unallocated to the connections of type p_0 or higher. This restriction does not apply to $p_0 = 1$, meaning that any unallocated part of the capacity can always be assigned to connections of type 1. Furthermore, one packet per frame normally accounts for a small percentage of the total link capacity. Therefore, the inequality does not necessitate any waste of capacity, nor does it impose any important restriction on capacity allocations.

It then follows that in the case where the arriving and departing frames are synchronous, a type- p packet which has arrived at a node n via some link l during an arriving type- p frame f will receive service before the type- p frame after f expires. It also follows that each connection k will maintain its original smoothness property throughout the network, i.e. for a connection of type- p with the allocated transmission rate r_k , the packet stream maintains the (r_k, T_p) smoothness throughout its path to the destination.

A node n'' for carrying out the stop-and-go queuing strategy with multiple frame sizes is illustrated in FIG 7. The node n'' is designed for the case where the arriving frames on all incoming links such as link l and the departing frames on all outgoing links such as link l' are synchronous for all frame sizes. The node n'' comprises a FIFO queue 50 for each frame size and

corresponding packet types $p=P, p=P-1, \dots, p=1$. If there are only two packet types, i.e. $p=2$ and $p=1$, then there are only two queues 50. Packets arriving on the link 1 are separated onto lines 53 according to packet type- p by the demultiplexer 52 and loaded into the appropriate FIFO queues 50.

In order to comply with the stop-and-go queuing rules A and B stated above, the service controller 54 operates as follows: At the beginning of each type- p frame, $p=1, 2, \dots, P$, the service controller 54 marks the present load in the type- p queue as eligible for service. At the beginning of each frame of type $p=P$, the service controller sets the switch 56 so that the server 58 starts serving eligible packets in the type- P queue and continues until the eligible type- P packets are served. In each type- $P-1$ frame (whose beginning always coincides with the beginning of one type P frame, see, e.g., FIG 6A and FIG 6B which show type-2 and type-1 frames), after the eligible load of queue P is served, transmission of the eligible load of queue $P-1$ is started. During each type $P-1$ frame, service of queue $P-1$ may several times be interrupted on a non-preemptive basis before its eligible load is transmitted. This is because each time a new type- P frame starts, new packets may become eligible at queue P . Finally, when the eligible load of queue $P-1$ is served, service of queue $P-2$ starts. This procedure

continues all the way down to queue 1. Once there is no eligible load at any queue, the server 58 is idle.

A typical pattern of service shifts between the queues for the case P-2 is illustrated in FIG 8. In FIG 8, the abscissa is a time axis which indicates the times occupied by the type-2 and type-1 frames. The ordinate in FIG 8 indicates the particular queue being served, i.e. queue p-1, queue p-2, or idle. Thus, as can be seen in FIG 8, the type-2 frames are shorter than the type-1 frames. Therefore, the type-2 packets are of higher priority than the type-1 packets. At the start of each type-2 frame, the eligible packet load of the type-2 queue is served. Only then does the server serve the eligible load of the type-1 queue. Once eligible load of the type-1 queue and the eligible load of the type-2 queue are served, the server is idle. Thus, in the type-2 frame $[2T_2, 3T_2]$, first the type-2 queue is served, then the type-1 queue is served and then the server is idle. Similarly, in the type-2 frame $[3T_2, 4T_2]$, the beginning of which coincides with the beginning of the type-1 frame $[T_1, 2T_1]$, first the type-2 queue is served and only then is the type-1 queue served. Note that service of the type-1 queue is not interrupted at the start of the type-2 frame $[4T_2, 5T_2]$ because there are no higher priority type-2 packets eligible for transmission at this time.

As indicated above in connection with FIG 6B, in the case where at a node, the frames arriving on all the

incoming links are synchronous, but the departing frames on the outgoing links are asynchronous with respect to the arriving frames, an additional delay θ_l^p may be incorporated into each incoming link l for the type- p packets to synchronize the arriving and departing frames of type- p . This can be accomplished by modifying the function of the service controller 57 of the link l as follows. At the beginning of each type- p frame $p=1,2,\dots,P$, the load present in the queue p is marked. This marked load is designated as eligible for transmission θ_l^p seconds later.

When the arriving and departing frames at a node are synchronous, the total buffer capacity per link l required to eliminate any chance of buffer overflow is

$$B_l = \sum 2C_l PT_p \quad (13)$$

This required buffer capacity is always less than $2C_l T_1$. When the arriving and departing frames are asynchronous, the required buffer size per link l can be as large as

$$B_l = \sum (2T_p + \theta_l^p) \cdot C_l^p \quad (14)$$

This is always smaller than $3T_1 C_l$.

The delay D^p associated with a type- p packet may be represented as follows

$$D^p = DP + d^p = \tau + Q^p + d^p \quad (15)$$

where DP is the total connection delay, d^p is the jitter delay, τ is the total end-to-end processing and

propagation delay, and Q_P is the connection queuing delay for a type-p packet.

Furthermore

$$-T_P < d_P < T_P \quad (16)$$

5 and

$$HT_P < Q_P < 2HT_P \quad (17)$$

In addition, it should be noted that in the multiframe case, bandwidth can be allocated to a connection of type-p in incremental steps of

$$\Delta r_P = L/T_P \text{ bits/sec} \quad (18)$$

In brief, the required buffer capacity per link for a type-p packet is proportional to T_P as is the queuing delay. On the other hand the incremental bandwidth allocation unit Δr_P is inversely proportional to T_P . However, while the queuing delay and incremental bandwidth are coupled together for each type of connection, it is now possible to provide small queuing delays for certain types of connections by choosing a small frame size while allocating bandwidth in fine segments for other types of connections by choosing a larger frame size.

As a numerical example, consider a network with link capacities of 150 Mbits/sec and packet size of $L=400$ bits. In addition, assume only two frame sizes, $T_1=16$ msec and $T_2=1$ msec, are employed. The incremental unit of bandwidth allocation for type-1 and type-2 connections will be $\Delta r^1=25$ Kbits/sec and $\Delta r^2=400$

kbits/sec. The maximum queuing delay for a type-1 connection with H=5 is $Q^1=160$ msec and the maximum queuing delay for a type-2 connection is $Q^2=10$ msec. This suggests that continuous bit stream oriented traffic with high vulnerability to end-to-end delays should be set up in the network as type-2 connections while low speed data communications services can be designated as type-1 traffic.

The benefits of freedom from congestion and bounded delay provided by the framing congestion control strategy are accomplished at the cost of the admission policy which enforces a smoothness property on packet arrivals, i.e. admitting only $r_k.T$ packets per connection k in each frame of size T . Since an averaging period of only a fraction of a second is insufficient to smooth out the statistical fluctuations of traffic sources, for practical purposes, the admissions policy requires that capacity be allocated based on the peak rates of connections. If the framing strategy is uniformly applied to all of the services in a broadband network, low utilization of transmission capacity may result. This problem can be alleviated by combining the framing congestion control strategy with other traffic management policies.

FIG 9 schematically illustrates a node n''' . The node n''' of FIG 9 is the same as the node n'' of FIG 7 except for the addition of the queue 100. A category of packets referred to herein as type $p=0$ packets is routed

by the multiplexer 52 into the queue 100. The service controller 54 controls the switch 56 in the same manner as is done in connection with FIG 7 except that the queue 100 for type-0 packets has replaced the idle position of the switch 56. Thus during any period of time that none of the queues 50 are being served, the server 58 is connected to queue 100.

This otherwise idle time for the server 58 is composed of two parts.

- 10 a) The fraction of the frames corresponding to any part of the transmission bandwidth not allocated to the traffic of type- $p \geq 1$, and
- b) The fraction of the bandwidth allocated to the traffic of type $p \geq 1$ which is being
- 15 underutilized by the corresponding sources.

It is important to notice that the new category of traffic may be regulated according to any desired policy or transmitted in any order, and can itself comprise different classes of connections, as well as datagrams, each being subject to a different congestion control policy. As long as the type -0 packets have low priority with respect to the eligible packets of type $p \geq 1$, their presence does not cause any disruption in the proper functioning of the framing strategy, as the type-0 packets are only served when the server would otherwise be idle. (In this case, the equality (12) above applies when $p_0 = 1, 2, \dots, P$)

20

25

In this integrated environment, the framing strategy is best suited for the periodic bit streams, real time traffic, and other services with stringent packet loss or delay requirements, while the bursty
5 services with less sensitivity to loss or delay should be included in the type-0 category.

In short, a framing congestion control strategy for a packet network which utilizes multiple frame sizes has been disclosed. The congestion control strategy
10 comprises an admissions policy for each connection at its source node and a stop-and-go queuing strategy at the network nodes. The multiple frame sizes enable queuing delays to be made small for some connections while enabling bandwidth to be allocated in small
15 increments for other connections.

Finally, the above-identified embodiments of the invention are intended to be illustrative only. Numerous alternative embodiments may be devised by those skilled in the art without departing from the spirit and
20 scope of the following claims.

CLAIMS

1. A method for transmitting packets in a packet switching network comprising
- for each connection k of a particular
- 5 type- p , $p=1, \dots, P$, limiting the number of packets admitted to the network in a frame of duration T_p to $r_k T_p$, where T_p is selected individually for each connection type- p and where r_k is a transmission rate allocated to connection k , and
- 10 at each node in said network, delaying a packet belonging to a connection of type- p and arriving at a node in said network in a particular arriving frame of duration T_p at least until the particular incoming frame has expired.
- 15 2. The method of claim 1 wherein said delaying step comprises servicing a packet belonging to a connection of type- p , which has arrived at a node on a particular incoming link during a particular arriving frame of duration T_p , during the first departing frame
- 20 of duration T_p on a particular outgoing link which starts after the end of the particular arriving frame.
3. A method for transmitting packets in a packet switching network comprising a plurality of nodes interconnected by links, each of said packets belonging
- 25 to a connection k of type p , $p=1, \dots, P$, said method comprising transmitting packets belonging to connections of type- p between nodes via said links in time frames of duration T_n , the duration T_n being chosen individually

for each connection type- p , the number of packets being transmitted by a connection k of type- p in one of said frames of duration T_p being bound by $r_k \cdot T_p$ for each link traversed by the connection k , r_k being a rate allocated to connection k .

4. The method of claim 3 wherein each packet of a connection k of type- p arriving at a node via an incoming link in a particular arriving frame of duration T_p and destined for a particular outgoing link is transmitted via said particular outgoing link in the first departing frame of duration T_p which starts at or after the end of the particular arriving frame.

5. The method of claim 4 wherein said arriving and departing frames are synchronous.

6. The method of claim 4 wherein said arriving and departing frames are asynchronous.

7. The method of claim 3 wherein the number of packets of connection k of type- p admitted to the network at a source node for connection k in a frame of duration T_p is limited by $r_k \cdot T_p$.

8. The method of claim 4 wherein T_p is an integral multiple T_{p+1} .

9. The method of claim 8 wherein a packet belonging to a connection of type- p has non-preemptive priority over a packet belonging to a connection of type $p-1$.

10. A method for transmitting packets in a packet switching network comprising a plurality of nodes

interconnected by links, each of said packets belonging to a connection k of type- p , $p=1, \dots, P$, said method comprising

- transmitting said packets between nodes
- 5 via said links in time frames, packets belonging to type- p connections being transmitted in frames of duration T_p , where T_p is chosen individually for each type- p , and
- delaying said packets at said nodes so
- 10 that each of said packets, belonging to a type- p connection and arriving at one of said nodes via an incoming link during a particular one of said frames of duration T_p and destined for a particular outgoing link, is transmitted from said one node via said particular
- 15 outgoing link in a frame of duration T_p which begins at or after the end of the frame of duration T_p in which the packet arrived at said one node.

11. The method of claim 10 wherein each connection k of type- p is allocated a rate r_k and the maximum
- 20 number of packets belonging to a connection k of type- p which can be transmitted in a frame of duration T_p is $r_k \cdot T_p$.

12. The method of claim 10 wherein each connection k of type- p is allocated a rate r_k and wherein the
- 25 maximum number of packets belonging to a connection k of type- p which can be admitted to the network in a frame of duration T_p is $r_k \cdot T_p$.

13. A node for use in a packet network for routing packets, each of said packets belonging to a connection k of type- p , $p=1, \dots, P$, said node comprising
- receiving means for receiving said packets via
- 5 one or more incoming links, said receiving means receiving said packets belonging to a connection k of type- p in arriving frames of duration T_p , where T_p is chosen individually for each connection type- p , and
- transmitting means for transmitting each of
- 10 said packets on a desired outgoing link, each packet belonging to a connection k of type - p being transmitted in the first departing frame of duration T_p on the desired outgoing link which starts at or after the end of the arriving frame of duration T_p in which the packet
- 15 arrived at the node.
14. The node of claim 12 wherein T_p is an integral multiple of T_{p+1} and wherein packets belonging to a connection of type- p have a non-preemptive priority over packets belonging to a connection of type- $p-1$ at said
- 20 node.
15. The node of claim 14 wherein
- said receiving means comprises
- demultiplexing means for separating packets belonging to connections of different types p , and buffer means
- 25 associated with each connection type- p for storing said packets, and
- said transmitting means comprises serving means for serving said buffer means so that packets

belonging to a connection of type-p have a non-preemptive priority over packets belonging to a connection of type p-1.

16. The node of claim 15 wherein said node includes additional buffer means for packets not belonging to a connection of one of said p types, said additional buffer means having a lower priority for service by said serving means than said buffer means associated with each connection type-p.

10 17. A method for transmitting packets in a packet switching network comprising a plurality of nodes interconnected by links, said method comprising transmitting packets belonging to different connection types between nodes in frames of different sizes and
15 delaying said packets at said nodes so that each packet arriving at one of said nodes via a particular incoming link in a particular arriving frame of a particular size is transmitted from said one node on a particular outgoing link in the first outgoing frame of said
20 particular size on said particular outgoing link which starts after said particular incoming frame expires.

1/7

FIG. 1

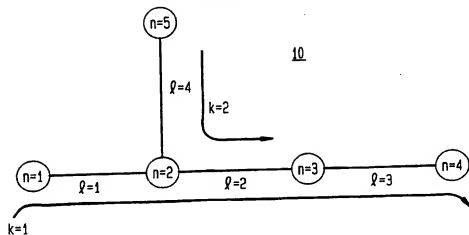


FIG. 2A

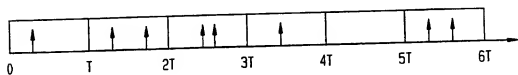


FIG. 2B

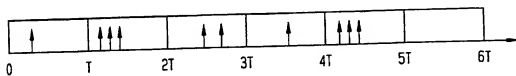


FIG. 3A

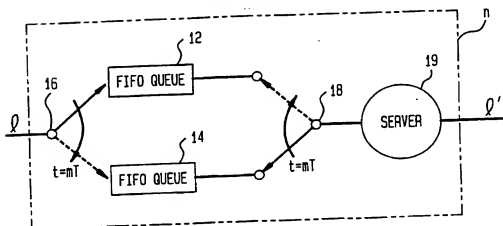


FIG. 3B

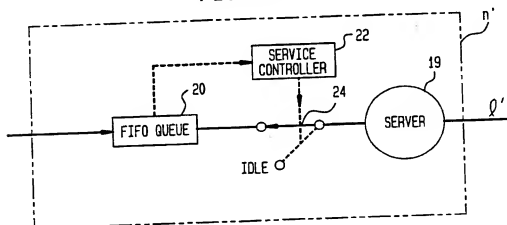


FIG. 4A

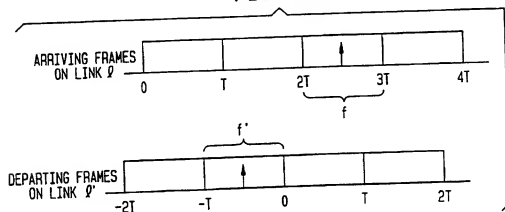


FIG. 4B

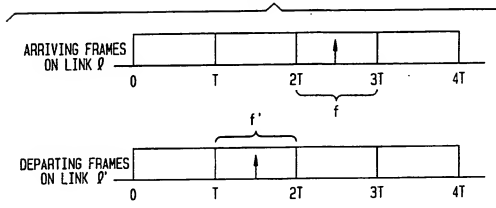


FIG. 5

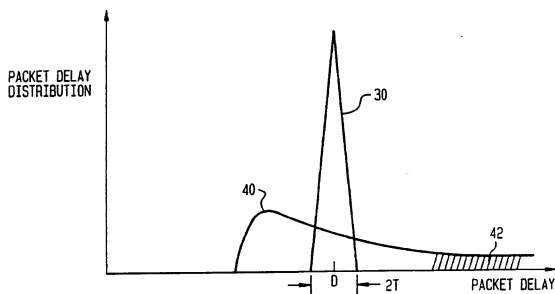
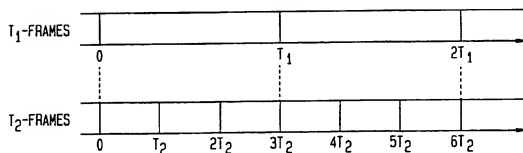


FIG. 6A



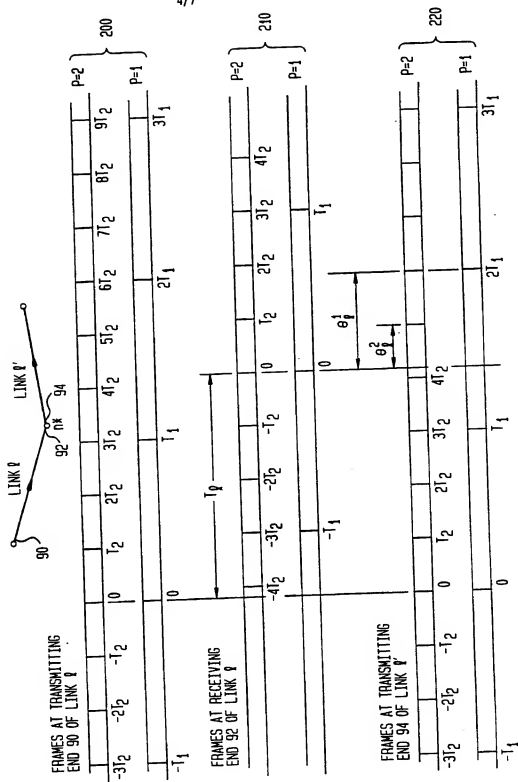
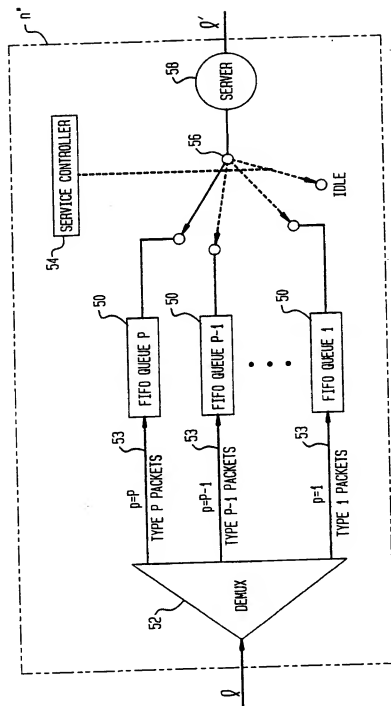


FIG. 7



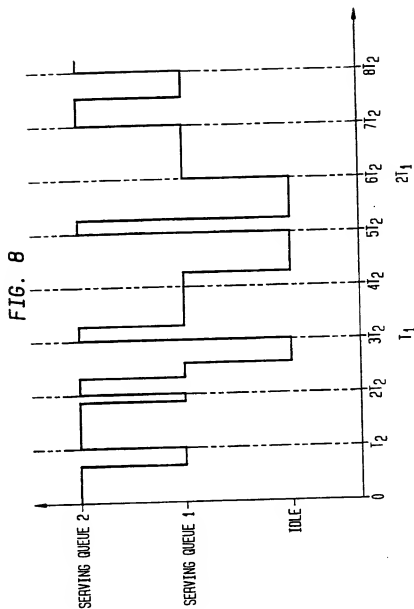
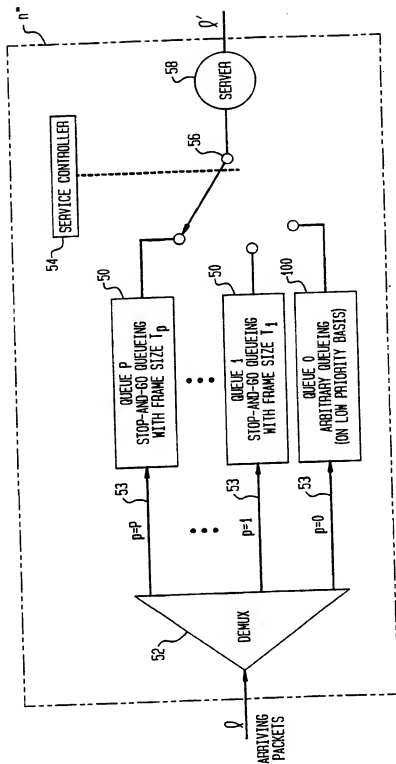


FIG. 9



INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US90/05621

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all *)

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): H04J 3/24; H04Q 11/04

US CL.: 370/60, 85.6, 94.1, 118; 340/825.51

II. FIELDS SEARCHED

Minimum Documentation Searched *

Classification System *

Classification Symbols

US 370/60, 77, 82, 84, 85.6, 85.12, 85.15, 94.1, 94.3, 100.1
105.1, 118 340/825/06, 825.5, 825.51, 825.52

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched *

III. DOCUMENTS CONSIDERED TO BE RELEVANT **

Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages **	Relevant to Claim No. **
Y	US, A, 4,858,232 (DIAZ et al.) 15 August 1989 See column 6, line 31 to column 7, line 12.	1-8,10-13.17
Y	IBM Technical Disclosure Bulletin Vol. 29 No. 9 See page 3756, line 32 to page 3757, line 5.	9,14-16
P,A	US, A, 4,914,650 (SRIRAM) 03 April 1990 See the entire document.	1-17
A	US, A, 4,878,218 (TAKADA) 31 October 1989 See the entire document.	1-17
A	US, A, 4,841,523 (ROFFINELLA et al.) 20 June 1989 See the entire document.	1-17
A	US, A, 4,779,267 (LIMB) 18 October 1988 See the entire document.	1-17
A	US, A, 4,734,908 (HEDLUND) 29 March 1988 See the entire document.	1-17

* Special categories of cited documents: **

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" documents referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is compared with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search *

Date of Mailing of this International Search Report *

18 DECEMBER 1990

11 FEB 1991

International Searching Authority *

Signature of Authorized Officer **

ISA/US

ALFUS H. HSU